

-1-

Date: 4-25-00 Express Mail Label No. EL 29 0724650 US

Inventors: William J. Dally and John W. Poulton  
Attorney's Docket No.: 2789.2004001

## TRANSITION-TIME CONTROL IN A HIGH-SPEED DATA TRANSMITTER

### RELATED APPLICATION(S)

This application claims the benefit of U.S. Provisional Application No.  
60/181,276, filed February 9, 2000, the entire teachings of which are incorporated  
5 herein by reference.

### BACKGROUND OF THE INVENTION

The transition time, rise time or fall time, of an output driver is the time required  
for the output signal to slew between two voltages, typically 20% and 80% of full  
swing. In certain prior-art systems, for example, transition time must be maintained  
10 larger than a specified minimum value to keep the derivative of the supply current, and  
hence the inductive switching noise (sometimes called simultaneous switching output  
(SSO) noise), within limits. At the same time, transition time must be kept smaller than  
a maximum value to avoid excessive delay of the signal.

Across variations in process parameters, supply voltage, and temperature,  
15 transition time can vary by a considerable amount, often by a factor of two or more. In  
some applications, the spread between maximum and minimum values for transition  
time is wide enough so that this variation is acceptable. In other applications, however,  
the window of legal transition times is small and such a large variation in transition time  
is unacceptable.

20 Transition time control has been employed in prior art systems with relatively  
slow signaling rates, where the bit time is greater than 10 gate delays. At such data  
rates, transition time can be controlled by using a tapped delay line to sequence the

stages of a segmented transmitter or by slowing the predriver stage of a transmitter. These techniques are discussed in Dally and Poulton, *Digital Systems Engineering*, Cambridge, pages 533-536.

Still other prior art systems have employed transition time control by controlling  
5 the transition time of a pre-driver which, in turn, controls the transition time of the output driver. The pre-driver transition time may be controlled by varying its supply voltage, controlling its supply current, or enabling a variable number of parallel pre-driver elements. As explained in Dally and Poulton, pp 533-536, slowing the transition  
10 time of the pre-driver in this manner can lead to severe inter-symbol interference, especially at high signaling rates. Because the output stage typically has significant gain, the predriver must be made very slow to give an output transition time that is a substantial fraction of a bit time. Often the pre-driver is so slow that it is not able to swing full-rail before the end of the bit time leading to significant inter-symbol interference due to the retained state.

15 In systems that operate at fast signaling rates, where the bit time is just a few loaded gate delays (less than 10), neither of these prior art transition control mechanisms is applicable. The transition time in such high-speed systems is just a few gate delays (less than 3) and thus comparable to the delay of a single tap of a tapped delay line. Because the entire transition must occur in just one (or at most two) taps of  
20 the delay line, it is not possible to smoothly sequence the transition by using a tapped delay line to sequence transmitter segments.

In such high-speed systems, the transition time is typically a large fraction of the bit time (usually 30%-50%). This is because a faster transition time would stress the bandwidth of the transmission medium (package, PC board, and connectors) without  
25 offering any substantial advantage. With such a ratio of transition time to bit time, it is not possible to control the transition time by slowing the predriver. To do so would require the predriver to have a delay much longer than the bit time and thus would cause intersymbol interference.

Because of these limitations, prior-art high-speed signaling systems have not employed transition time control and, as a result, have incurred large variations in transition time across process, voltage, and temperature corners.

#### SUMMARY OF THE INVENTION

5 In accordance with one aspect of the invention, a data transmitter comprises a data input and plural delay elements. The delay elements apply different delays to the data input in parallel to provide plural delayed data signals. A data output combines the delay and data signals so that a transition time of the data output is determined by difference in delays applied to the data input.

10 Prior art transition control systems controlled the transition time to be a fixed value, regardless of the bit time of the system. With a fixed transition time, a signaling system operating at a lower speed is forced to use a transition time optimized for the highest possible speed of operation, unduly stressing the bandwidth of the transmission medium.

15 In accordance with another aspect of the invention, transition time control controls the transition time of a controlled data signal to be proportional to bit time of a bit clock.

A clock signal may be applied to the delay elements, with different delays being applied to the data input by clocking the data input with different delayed clock signals.

20 Alternatively, the data input may be applied in parallel directly to the delay element. In either case, the delayed data signals are applied to plural driver circuits. Preferably, each delay element comprises CMOS inverters and the delay of the delay elements is determined by load capacitance.

25 Supply voltage to the delay elements may be controlled to control delay of the delay elements. In one embodiment, a circuit to control the supply voltage to the delay elements comprises first and second delay elements, each receiving a common clock signal. A phase comparator compares outputs of the first and second delay elements and controls a supply voltage applied to the first and second delay elements to control

phase difference of the outputs. Each of the first and second delay elements may comprise a sequence of  $n$  elements and the clock signal frequency may then be  $1/n$  times bit rate. The supply voltage may thus be varied to compensate for environmental changes in delay.

## 5 BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings in which like reference characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

Figure 1 illustrates an embodiment of the invention in which a clock signal is applied to delay elements and the data input is clocked with different delayed clock signals.

15 Figure 2 is a timing chart for the circuit of Figure 1.

Figure 3 illustrates an embodiment of the invention in which the input data is delayed directly in parallel delay elements.

Figure 4 is a timing chart for the circuit of Figure 3.

Figure 5 illustrates an array of delay elements for use in either Figure 1 or Figure 3.

Figure 6 illustrates a circuit to control the supply voltage to the delay element of Figure 5.

Figure 7 illustrates a modification of the circuit of Figure 6 for operation with a slow multiplexing clock.

25 Figure 8 illustrates an array of delay elements including serial/parallel connections of delay elements.

## DETAILED DESCRIPTION OF THE INVENTION

A description of preferred embodiments of the invention follows.

In a high-speed transmission system, where the bit time is less than 4 gate delays, prior art approaches to controlling rise time do not apply. A tapped delay line  
5 cannot be used since the desired transition time is comparable to the delay of a single tap. Slowing the predriver is also not appropriate as it will result in considerable ISI as the slow predriver stage will not reach a steady state before the end of each bit cycle.

The present invention overcomes these limitations and controls the rise time of a high-speed transmitter by segmenting the transmitter and driving each segment with a  
10 variable delay element driven from a common clock node. By appropriately adjusting the delays of the variable delay elements the segment switching times can be set at intervals that are a small fraction of a gate delay resulting in a controlled transition time comparable to a single gate delay. With this approach the timing resolution is set by the difference between element delays rather than by the delay of a single element. This  
15 gives a granularity of timing control fine enough to handle the fastest signaling systems.

In high-speed signaling systems it is advantageous to control the transition time to be not a fixed interval, but rather a fraction of the bit-time (e.g., 40%). With this approach, a signaling system operating at a lower speed (with a longer bit time) would use a proportionally longer transition time. Hence it requires less bandwidth out of the  
20 transmission medium and can use less expensive materials and components in constructing the transmission system. At very low signaling rates, of course, the transition time is maintained less than a fixed maximum to avoid noise problems that occur with very slow transition times.

The present invention achieves this variable bandwidth advantage by controlling  
25 transition time to be a fraction of the bit time. This is accomplished by adjusting the variable delay elements so the difference in delay between the slowest element and fastest element is equal to the desired fraction of a clock cycle.

A block diagram of an embodiment of the present invention is illustrated in Figure 1. The figure illustrates a segmented output driver with transition control. The

driver accepts data on line 101 and an input clock on line 102. The input clock is delayed by four delay elements 121-124 with delays  $t_{d1}$  to  $t_{d4}$  to generate sequencing clocks c1 to c4 on lines 103-106. These sequencing clocks are used to clock the input data into four flip-flops 125-128. The outputs of the flip-flops, d1-d4 on lines 107-110, are input to current drivers 129-132. The current drivers sum their currents onto output line 111 so that the current waveform on this line is the superposition of the currents from the four current drivers.

The block diagram of Figure 1 is best understood with reference to the waveforms of Figure 2. The top trace shows the data input line 101 which rises at the beginning of the trace and remains high during the valid window when it is being sampled by sequencing clocks c1 to c4. One skilled in the art of digital design will understand that the data input signal is preconditioned using latches to guarantee that it is stable during this valid window. The second trace shows the input clock, ck on line 102. The next four traces show the sequencing clocks c1 to c4 on lines 103-106. The delay of each element is slightly different with delay element 121 having the smallest delay and delay element 124 having the largest delay. The delay increases by a fixed amount per element to give four evenly spaced sequencing clocks. The figure illustrates how these parallel delay elements can generate sequencing clocks with a spacing,  $\Delta t_d$ , that is much less than the delay of the fastest element,  $t_{d1}$ . The spacing is set by the difference between the delay of two elements,  $\Delta t_d = t_{d2} - t_{d1}$ . This is in contrast to prior art transition time control systems based on tapped delay lines where the spacing of sequencing clocks must be at least as large as the delay of an element,  $t_d$ .

The next four traces, traces 7 through 10, show the outputs d1'-d4' of the individual current drivers 129-132 before they are summed on the line. For clarity in the figure we have shown these signals with an unrealistically short delay from the clock inputs of the flip flops to the corresponding outputs of the current drivers (e.g., from c1 on line 103 to the output of current driver 129). In practice there would be a much larger delay between these two signals. However the causality of the signals is easier to appreciate with the waveforms as drawn in Figure 2.

The rise time of an individual current driver,  $t_{r1}$ , is designed to be comparable to the spacing of the sequencing clocks,  $\Delta t_d$ , to ensure a smooth transition of the summed signal. The final output of the driver, the summed signal on data output line 111 is shown in the final trace. It has a rise time that is equal to  $3\Delta t_d + t_{r1}$ .

5        Figure 3 shows a block diagram for an alternate embodiment of the present invention in which the data, rather than the clock is delayed by a parallel arrangement of delay elements. In this embodiment the data in signal on line 101 is aligned with the clock, ck on line 102, by flip-flop 142. The aligned data signal, d0 on line 141, is then input to the four delay elements 121-124 with delays  $t_{d1}$  to  $t_{d4}$ . In this case, the delay  
10    elements directly generate the skewed data signals d1 to d4 on lines 107 to 110. As with the system of Figure 1, these data signals are then input to current drivers 129 to 132 which sum their outputs on data output line 111.

The embodiment of Figure 3 is advantageous in that it requires fewer flip-flops than the embodiment of Figure 1 and, thus, reduces clock loading. The embodiment of  
15    Figure 1 is preferred, however, in cases where the sequencing clocks c1 through c4 can be shared across multiple output drivers.

The embodiment of Figure 3 can be better understood by reference to the waveforms of Figure 4. The first trace shows the data input, din on line 101, and the second trace shows the clock, ck on line 102. Because din is sampled by only a single  
20    clock, ck on line 102, it need only be valid during a small timing window, as illustrated, about the rising edge of the clock to account for setup and hold time. This is in contrast to the wide timing window required for din in the embodiment of Figures 1 and 2.

The third trace shows the aligned data out of flip-flop 142, d0 on line 141. This is a version of the data signal aligned to the clock. As in Figure 3, we have purposely  
25    shown the clock-to-Q delay of the flip-flop much shorter than is realistic to improve the clarity of the figure. In reality there would be a much longer delay between the rising edge of clock and the transition on d0.

The next four traces show the outputs of the four delay elements, d1 to d4 on lines 107 to 110. In this figure we show the inputs to the current sources while in

Figure 2 traces with the same labels showed the outputs of the current sources with slower rise times. These traces illustrate how the parallel combination of delay lines is able to sequence signals with time differences,  $\Delta t_d$ , that are substantially smaller than the minimum delay of a delay element,  $t_{d1}$ .

- 5        The final trace of Figure 4 shows the data output signal. This is the superposition of the outputs of current drivers 129 through 132. As with the embodiment of Figures 1 and 2, the rise time of this signal is equal to  $3\Delta t_d + t_{r1}$ .

- One skilled in the art will understand that a high-speed driver with transition time control can be realized with many variations on the block diagrams of Figures 1 and 3. For example the driver may have a greater or smaller number of segments than the four segments shown in Figure 1. The output drivers may be voltage mode rather than current mode. Also, the drivers may be differential rather than single ended. The flip flops of Figure 1 may be replaced by latches, multiplexers, or a combination of latches, flip-flops, and multiplexers that aligns the data with the sequencing clocks.
- 10        Finally, the sequencing clocks may be generated with a combination of series and parallel delay elements or with such elements in combination with a multi-phase clock or a clock generated by an array oscillator.

- Figure 8 illustrates an alternative array of delay elements in which parallel delays are obtained through serial/parallel connections of delay elements. In particular, of four delays, two are provided by connecting the parallel delays  $t_{d1}$  and  $t_{d2}$  in series with a common delay element  $t_{d3}$ .
- 20

- One embodiment of the array of delay elements 121-124 of Figures 1 and 3 is illustrated in Figure 5. Each delay element comprises a pair of inverters. Other than the first element, each element also includes a capacitor to increase the delay of the element. For example, delay element 122 comprises inverters 153 and 160, and the output of inverter 153 is loaded by capacitor 156 with a capacitance of C. Subsequent delay elements use proportionally larger capacitors. Delay element 123 has a capacitor 157 with value 2C, and element 124 has capacitor 158 with value 3C.
- 25



The delay of a CMOS inverter increases in proportion to its output capacitance according to the formula,  $t_d = t_{d0} + C t_c$ . In this formula  $t_{d0}$  is the delay of an inverter with no output load and  $t_c$  is the increase in inverter delay per unit of output load capacitance. Thus, the delay of element 121 in Figure 5 is  $t_{d1} = t_{d0} + C_p t_c$  where  $C_p$  is the parasitic element of the intermediate node of the delay element. The delay of element 122 is  $t_{d2} = t_{d0} + (C_p + C) t_c = t_{d1} + C t_c$ . Delay element 123 has delay  $t_{d3} = t_{d0} + (C_p + 2C) t_c = t_{d2} + C t_c$ , and delay element 124 has delay  $t_{d4} = t_{d0} + (C_p + 3C) t_c = t_{d3} + C t_c$ . Because the capacitance is increased by a fixed amount,  $C$ , at each stage, the delay also increases by a fixed amount,  $\Delta t_d = C t_c$  at each stage. The capacitance,  $C$ , is chosen so the increment in delay,  $\Delta t_d = C t_c$ , is the required fraction of the bit time.

To compensate for the variation in delay due to process, voltage, and temperature variation the delay of inverters 152-155 can be varied by varying the supply voltage of each inverter. The supply voltage of these inverters is separated from the main supply and tied to control voltage,  $v_{ctrl}$  on line 151, to facilitate this compensation. As will be explained below, this control voltage can also be used to make the variation in delay between elements, and hence the transition time of the output, proportional to the bit time.

One skilled in the art will understand that the delay elements 121-124 can be implemented in many ways. Element delay can be varied by varying the drive of each inverter rather than varying the capacitive load. Delay can also be varied by varying the current to each stage. A differential delay element can be used rather than a single-ended element. A different circuit topology, for example a source-coupled FET delay stage can be used in place of a CMOS inverter.

A circuit that both compensates the delay elements of Figure 5 for process, voltage, and temperature variations and at the same time adjusts the transition time to be proportional to bit time is illustrated in Figure 6. The circuit comprises two delay elements, 176 and 177, of the same type used in Figure 5, a phase comparator and charge pump, 170, and a voltage follower, 180. Delay element 176 has no additional capacitance while delay element 177 is loaded with a capacitor with capacitance  $mC$ .

The phase comparator and charge pump may be of any type. The preferred embodiment uses a combined phase comparator and charge pump as disclosed in pending patent U.S. patent application no. 09/414,761, filed October 7, 1999, by Dally et al. for Combined Phase Comparator and Charge Pump Circuit.

5 This circuit uses feedback to adjust the control voltage,  $v_{ctrl}$  on line 151, so that the difference in delay between delay element 176 and delay element 177 is exactly one cycle of clock  $ck$ , i.e., one bit clock cycle,  $t_{bit}$ . If the difference between the delay of the elements is less than  $t_{bit}$ , the rising edge of the delayed clock signal  $c_{xm}$  on line 179 will lead the rising edge (of the subsequent clock cycle) of the delayed clock signal  $c_{x1}$  on  
 10 line 178 and the charge pump will decrease the control voltage. Decreasing the control voltage increases both the overall delay of both delay elements and the difference in delay between the elements. Eventually, this feedback will bring the two clock edges into alignment at the point where the difference in delay is exactly equal to  $t_{bit}$ . Similarly, if the delay difference is greater than  $t_{bit}$ , signal  $c_{x1}$  will lead signal  $c_{xm}$  and  
 15 the charge pump will increase  $v_{ctrl}$  to reduce the delay and again bring the two signals into alignment.

When the control loop has converged, the transition time of the driver is set to  $(4/m)t_{bit}$ . Because delay element 176 has delay  $t_{dc1} = t_{d0} + C_p t_c$  and delay element 177 has delay  $t_{dc2} = t_{d0} + (C_p + mC)t_c$ , the difference in delay between the two elements is  
 20  $\Delta t_{dc} = mC t_c$ . When the loop has converged, this difference is equal to the bit time:  $\Delta t_{dc} = t_{bit}$ . Thus  $t_{bit} = mC t_c$  and we have  $\Delta t_d = C t_c = t_{bit}/m$ . Since the drivers of Figures 1 and 3 have a rise time of  $t_r = 3\Delta t_d + t_{r1}$  or if  $t_{r1} \sim \Delta t_d$ ,  $t_r \sim 4\Delta t_d = (4/m)t_{bit}$ . In the case where  $m=10$ , for example, the transition time is controlled to be 40% of the rise time.

To prevent the transition time from exceeding a maximum value when the bit  
 25 clock,  $ck$ , is run very slowly, a diode clamp 190 is placed on the output of the charge pump that prevents  $v_{ctrl}$  from being decreased below a minimum value. This limits the rise time to be no greater than the delay corresponding to this clamped value. The diode clamp may be implemented with a diode-connected MOSFET.

In a multiplexing transmitter, the fastest available clock signal has a period that is  $t_{ck} = nt_{bit}$  where  $n$  is the multiplexing factor, typically between 2 and 20. The compensation circuit of Figure 6 can be modified to operate with such a slow multiplexing clock by placing multiple delay elements in series as illustrated in Figure 7  
5 for the case where  $n=2$ . In this circuit, the upper path of identical delay elements 176 and 186 has a delay that is twice the delay of element 176. Similarly, the lower path of identical delay elements 177 and 187 has a delay that is twice the delay of element 177. Thus when the feedback brings nodes  $cyl$  on line 188 and  $cym$  on line 189 into phase we have  $2\Delta t_{dc} = t_{ck} = nt_{bit}$  or for the case where  $n=2$ ,  $\Delta t_{dc} = t_{bit}$  as desired. One skilled in  
10 the art will understand that a clock at  $n$  times the bit rate can be accommodated by placing  $n$  copies of delay element 176 in series on the upper clock path and  $n$  copies of delay element 177 in series on the lower clock path.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that  
15 various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.